July 19, 2023

Voluntary Commitments by Microsoft to Advance Responsible AI Innovation

Commitments at-a-glance

Building safe AI systems based on robust evaluation, verification, and validation.

1. Microsoft will continue to test its AI systems prior to release and on an ongoing basis using red-teaming and systematic measurement techniques. For high-risk systems, Microsoft will commit that it will ensure red teaming is conducted before deployment by qualified experts that are independent of the product teams building those systems and will share summaries of that testing with key stakeholders as appropriate.
2. Microsoft will participate with fellow leading AI developers in a focused forum to develop evaluation standards for emerging safety and security issues and will otherwise contribute to the development of ecosystem functions to enhance the safety, security, and transparency of AI systems.
3. Microsoft will deploy new state-of-the-art provenance tools to help the public identify AI-generated audio-visual content and understand its provenance.
4. Microsoft will implement the NIST AI Risk Management Framework and attest to alignment with it to customers.
5. Microsoft will implement robust reliability and safety practices for its high-risk models and applications, ensuring a layered safety-by-design approach so that models and applications remain safe, secure, and within human control.

Securing the use of Microsoft AI systems for highly capable models.

6. Microsoft will ensure that the cybersecurity risks of our AI products and services are identified and mitigated as part of our overall approach to responsible development and deployment.
7. Microsoft will participate in an approved multistakeholder exchange of information about critical safety and security threats to highly capable models.
8. Microsoft will support the development of a licensing regime to regulate the secure development and deployment of highly capable models.
9. Microsoft will support the development of an expanded 'know-your-customer' concept for AI services, building upon the same concept that has been implemented with respect to high-risk financial services.

Increasing the trustworthiness of Microsoft's AI systems.

10. Microsoft will commit that it will continue to ensure that our AI systems are designed to inform the public when it is interacting with an AI system and that the system's capabilities and limitations are communicated clearly, including via model- and application-level documentation.
11. Microsoft will increase investment in its academic research programs to ensure researchers outside Microsoft can access Microsoft's foundation models and the Azure OpenAI Service to undertake research and validate findings.

12. Microsoft will collaborate with the National Science Foundation to explore Microsoft's participation in a pilot project to inform efforts to stand up the National AI Research Resource, including by facilitating independent academic research relating to the safety of AI systems.
13. Microsoft will release an annual transparency report to inform the public about its policies, systems, progress, and performance in managing AI responsibly and safely.
14. Microsoft will support the development of a national registry of high-risk AI systems that is open for inspection so that members of the public can learn where and how those systems are in use.

Further Detail on the Commitments

Building safe AI systems based on robust evaluation, verification, and validation.

1. **Microsoft will continue to test its AI systems prior to release and on an ongoing basis using red teaming and systematic measurement techniques. For high-risk systems, Microsoft will commit that it will ensure red teaming is conducted before deployment by qualified experts that are independent of the product teams building those systems and will share summaries of that testing with key stakeholders as appropriate.**

   Microsoft's Responsible AI Standard guides engineering teams to identify potential harms, measure their propensity to occur, and build mitigations to address them. At Microsoft, we have further developed red teaming techniques, which were originally developed to identify cybersecurity vulnerabilities, to stress test AI systems using multi-disciplinary teams with a wide range of expertise, including privacy, security, and fairness. A summary of some of our learning about red teaming large language models is available here.

   We can build on this practice by committing that for OpenAI foundation models that we make available via the Azure OpenAI Service, we will refer our customers to the red teaming evaluation results that OpenAI makes available. For new and highly capable foundation models that Microsoft may train itself in the future, we will organize red team testing by one or more independent experts, external to Microsoft, prior to release of those models. The topics covered by such red teaming will align with industry best practice and include testing of dangerous capabilities. We will share summarized results of such testing with key stakeholders in a manner that is appropriate given safety and security considerations associated with their release.

At the application level, for the deployment of high-risk systems that Microsoft makes available, we will choose red team testers who are independent of the product teams that are building these systems, adopting a best practice from the financial services industry. We will rely upon these red team testers, together with our product teams who are responsible for systematic evaluations of the products that they build, to help us identify, measure, and mitigate potential harms. Furthermore, to continually monitor, track, and evaluate our AI systems, we will implement systematic measurements using responsible AI metrics to understand new issues specific to generative AI experiences, such as the extent to which a model's output is supported by information contained in input sources. (We announced the first of these metrics as part of our Azure OpenAI Service at Build, our annual developer conference.)

2. **Microsoft will participate with fellow leading AI developers in a focused forum to develop evaluation standards for emerging safety and security issues, and otherwise contribute to the development of ecosystem functions to enhance the safety, security, and transparency of AI systems.**

   Microsoft, together with Alphabet, Anthropic, and OpenAI, each commit to the safe and responsible development and deployment of AI, and to the protection of public safety in relation to emerging AI technologies. We will:

   - Work with the White House and other federal agencies to identify or establish a focused forum to develop recommendations for standards focused on evaluating emerging safety and security issues and mitigating potential large-scale risks posed by AI systems. These recommended standards will be clear, definite, and capable of being evaluated independently by third parties.

   - At the end of 2023, provide the White House with an update on progress towards recommended standards, and our intended sustained efforts going forward.

   In addition, and as noted throughout this paper, Microsoft is committed to playing its role to support the development of the necessary ecosystem functions to ensure that there are the providers, standards, and practices in place to help enhance the safety, security, and transparency of AI systems. This includes supporting the development of the audit and assurance ecosystem, test and evaluation professionals and tooling, undertaking research on safe and responsible practices, and participating in standards development organizations.

3. **Microsoft will deploy new state-of-the-art provenance tools to help the public identify AI-generated audio-visual content and understand its provenance.**

   At our annual developer conference, Build, Microsoft announced new media provenance capabilities. Using the C2PA specification from the Coalition for Content Provenance and Authenticity, we will mark and sign AI-generated images from Microsoft Designer and Bing Image Creator with metadata about their origin, enabling users to verify that images from those services were generated by AI. This marking will happen automatically, as part of the image generation process. We expect that the engineering of this system will be complete by the end of 2023.

Microsoft co-founded the Coalition for Content Provenance and Authenticity following our earlier provenance initiative Project Origin and our foundational building and prototyping work for provenance technologies. You can read more about our long-standing provenance efforts here.

4. **Microsoft will implement the NIST AI Risk Management Framework and attest to alignment with it to customers.**

The NIST AI Risk Management Framework (AI RMF) is a strong, existing framework for the U.S. Government that we believe provides an important foundation for the United States and world to build upon. It was developed in a consensus-driven and transparent manner, and it has been acknowledged by governments, civil society organizations, international institutions, and key industry players as a valuable resource. The AI RMF, in short, establishes a durable approach that can keep pace with developments in technology and responsible AI practice.

Microsoft participated in the development of the AI RMF and our internal Responsible AI Standard is closely aligned with it. We will commit to moving forward to ensure implementation of the AI RMF across Microsoft, and we will attest to alignment with the AI RMF in our contracts with customers.

We continue to believe that there is substantial merit in using federal procurement mechanisms to accelerate adoption of the AI RMF. This would enable the U.S. Government to capitalize on existing momentum around the AI RMF and implement an effective, initial regulatory response that can evolve as the regulatory landscape matures. More specifically, the following steps could be considered as part of a comprehensive approach to implementing the AI RMF using an Executive Order:

- **Require self-attestation by vendors of AI RMF alignment.** Self-attestation is used by the government to advance cybersecurity standards amongst federal suppliers. A similar mechanism can be applied to the AI RMF. The Office of Management and Budget (OMB) could issue guidance requiring federal agencies procuring AI services for use in critical decision systems to only do so from suppliers that have self-attested that they meet a minimum bar for implementation for the AI RMF. The minimum bar could be set by the NIST AI RMF Program Office mentioned below.

- **Establish a NIST AI RMF Program Office to advance coordination and enablement**. We suggest the creation of an NIST AI RMF Program Office to provide ongoing guidance for the Framework and promote adoption of it across agencies. This Program Office could also work with the new "Agency Equity Teams" required by EO 14091 on Advancing Racial Equity, to include guidance that helps small- and medium-sized organizations.

- **Develop responsible procurement resources.** The General Services Administration (GSA) and OMB could be directed to develop voluntary, standard contract language for agencies that are procuring critical decision systems, obligating a baseline set of actions in line with the Framework's recommendations.

  Additionally, NIST's important work to build out AI RMF "Profiles" (guides on how the AI RMF applies to specific sectors and/or systems) could include the development of specific profiles for public sector uses of critical decision systems.

- **Advance training and education.** The NIST AI RMF Program Office, coupled with GSA and Agency Equity Teams, could deliver training on AI trustworthiness for individuals responsible for

acquiring or procuring critical decision systems. This would support acquisition professionals in important roles that define the scope of contract solicitations, set contract requirements, or make vendor determinations. Training would cover the technology's risks and benefits in order to help acquisition professionals determine whether the software under consideration meets standards for performance and does not unlawfully discriminate.

- **Augment baseline AI governance requirements for agencies.** Federal agencies could be required to implement the NIST AI RMF in their own AI development. In time, this could be supplemented with mandatory responsible AI controls for government systems.

5. **Microsoft will implement robust reliability and safety practices for high-risk models and applications, ensuring a layered safety-by-design approach so that models and applications remain safe, secure, and within human control.**

   For highly capable models that Microsoft builds, we will produce a safety plan ahead of commencing model training. This safety plan will outline the mitigations we will implement to ensure safety confidence throughout the training process.

   At the application level, for high-risk AI systems, Microsoft will:

   - Evaluate the operational factors and ranges within which those AI systems are expected to perform reliably and safely, remediate issues, and provide related information to our customers;
   - Design those systems to minimize the time to remediation of predictable or known failures; and
   - Ensure that those systems are subject to ongoing monitoring, feedback, and evaluation so that we and our customers can identify and review new uses, identify and troubleshoot issues, manage and maintain the systems, and improve them over time.

   Where Microsoft is building an AI system that uses a highly capable model to control critical infrastructure, we will implement a layered approach to system safety, ensuring that the system is subject to "safety brakes" that ensure that the AI system remains within human control at all times. While the specific implementation of "safety brakes" will vary across different systems, a core design principle that Microsoft will adopt is that the system should possess the ability to detect and avoid unintended consequences, and it must have the ability to disengage or deactivate in the event that it demonstrates unintended behavior. Microsoft will also ensure that such systems embody best practice in human-computer interaction design.

Securing the use of Microsoft AI systems for highly capable models.

6. **Microsoft will ensure that the cybersecurity risks of our AI products and services are identified and mitigated as part of our overall approach to responsible development and deployment.**

   Microsoft makes the following voluntary AI cybersecurity commitments, which we have grouped by NIST Cybersecurity Framework 2.0 (Draft) functions. These specific commitments build upon the commitments made elsewhere in this paper, in some cases enlarging upon them to address core

cybersecurity requirements. For ease of reference, we have italicized commitments made elsewhere in this paper, so that the list below forms a comprehensive list of our AI cybersecurity commitments.

Govern

- *Microsoft will implement the NIST AI Risk Management Framework and attest to alignment with it to customers.*
- *Microsoft will support the development of a licensing regime to regulate the secure development and deployment of highly capable models.*
- Microsoft will participate with governments, standards bodies, and fellow leading AI developers in defining and evolving requirements and standards for AI cybersecurity, including those supporting any future licensing or regulatory regimes.
- Microsoft will link our Security Development Lifecycle (SDL) with our Responsible AI Governance Framework to ensure that cybersecurity risks inform AI risk management and that AI risks inform security development.

Identify

- *Microsoft will support the development of a national registry of high-risk AI systems that is open for inspection so that members of the public can learn where and how those systems are in use.*
- Microsoft will maintain an internal inventory of AI systems to support our Security Development Lifecycle (SDL) and Responsible AI Governance.
- Microsoft will contribute to the development of, and consider adopting where appropriate, standards for bills of materials (BOMs) within AI systems and for content provenance and authenticity.

Protect

- *Microsoft will continue to test its AI systems prior to release and on an ongoing basis using red teaming and systematic measurement techniques. For high-risk systems, Microsoft will commit that it will ensure red teaming is conducted before deployment by qualified experts that are independent of the product teams building those systems and will share summaries of that testing with key stakeholders as appropriate.*
- *Microsoft will participate with fellow leading AI developers in a focused forum to develop evaluation standards for emerging safety and security issues and will otherwise contribute to the development of ecosystem functions to enhance the safety, security, and transparency of AI systems.*
- *Microsoft will deploy new state-of-the-art provenance tools to help the public identify AI generated audio-visual content and understand its provenance.*
- Microsoft will continue to employ strong identity and access control, holistic security monitoring (for both external and internal threats) with rapid incident response, and continuous security validation (such as simulated attack path analysis) for our AI environments. Model weights will be encrypted-at-rest and encrypted-in-transit to mitigate potential model theft where applicable. Additional or more stringent security controls will be applied based on risk such as when protecting highly capable models.

- Microsoft will incorporate AI-specific guidance into our Security Development Lifecycle (SDL) and, where appropriate, publish that guidance for review and as a resource for other AI system developers.
    - Microsoft's SDL is mapped to NIST SP 800-218 (Secure Software Development Framework) and to the requirements arising from the Biden Administration's Executive Order 14028 (Improving the Nation's Cybersecurity).
- Microsoft will assist in and encourage the creation and adoption of testing frameworks to advance effective testing of AI systems for properties such as security.

Detect

- Microsoft will continually monitor, track, and evaluate the cybersecurity of our AI systems, and we will use metrics to measure and understand systemic issues. This monitoring will include detection of unauthorized data access or exfiltration and modifications to security configuration from either insider or external actors. We will use threat intelligence, including that provided by the Federal government through briefings and threat sharing programs, to inform our cybersecurity program. Microsoft has also established a bug bounty program.
- Microsoft will perform periodic security validation works to continuously validate our detection efficacy and make improvements where appropriate.
- Microsoft will - as with all our platforms - invest in preventing, detecting, and disrupting abusive and malicious use of our AI infrastructure, technologies, and products. We will use a combination of process, technical, and legal controls and review their effectiveness, improving them as needed.

Respond

- *Microsoft will release an annual transparency report to inform the public about its policies, systems, progress, and performance in managing AI responsibly and safely.*
- Microsoft will include in its AI transparency report, as appropriate, cybersecurity measures and information.
- Microsoft will disclose vulnerabilities related to our AI technologies in accordance with our CVE process (consistent with the CVE Program's guidance) and will follow coordinated vulnerability disclosure when dealing with supply chain vulnerabilities. Microsoft will not release vulnerability information publicly that may place customers, suppliers, partners, or competitors in imminent danger.
- Microsoft will publish an update to the AI/ML Pivots to the Security Development Lifecycle Bug Bar, which provides AI/ML-specific guidance for triaging security issues.
- Microsoft will continue to investigate and respond to insider threats in accordance with our internal policies and procedures and will work with Law Enforcement as appropriate.

7. **Microsoft will participate in an approved multistakeholder exchange of information about critical safety and security threats to highly capable models.**

   Microsoft recognizes that safe and secure development and deployment of highly capable models will require the timely exchange of information between the private and public sectors about critical safety and security threats and mitigation techniques. Microsoft is supportive of exploring appropriate models to pursue that type of information exchange, utilizing cybersecurity and digital safety best practices, and building upon existing channels of communication as appropriate.

8. **Microsoft will support the development of a licensing regime to regulate the secure development and deployment of highly capable models.**

   To secure the beneficial use of highly capable AI models and avoid their proliferation into the hands of bad actors, Microsoft is supportive of licensing regulations that would impose requirements on model developers and AI datacenter providers. We also support the establishment of a new regulator to bring this licensing regime to life and oversee its implementation.

   In our view, a licensing regime for highly capable AI models should be designed to fulfil three key goals. First and foremost, it must ensure that safety and security objectives are achieved in the development and deployment of highly capable AI models. Second, it must establish a framework for close coordination and information flows between licensees and their regulator, to ensure that developments material to the achievement of safety and security objectives are shared and acted on in a timely fashion. Third, it must provide a footing for international cooperation between countries with shared safety and security goals, as domestic initiatives alone will not be sufficient to secure the beneficial uses of highly capable AI models and guard against their misuse. We need to proceed with an understanding that it is currently trivial to move model weights across borders, allowing those with access to the 'crown jewels' of highly capable AI models to move those models from country to country with ease.

   More specifically, we believe the broad parameters of such a licensing regime ought to be as follows:

   - An initial key task for the administrator of the licensing regime will be to define the regulatory threshold for the regime. From Microsoft's perspective, the objective is not to regulate the rich ecosystem of AI models that exists today and should be supported into the future. Instead, the focus should be on the small number of AI models that are very advanced in their capabilities or that redefine the frontier. A capability- or compute-based threshold strikes us at the most appropriate approach, with a capability-based threshold likely to be more durable over time.

   - Above the regulatory threshold, developers of highly capable AI models should be required to provide advance notification of large training runs, undertake comprehensive risk assessments focused on identifying dangerous or breakthrough capabilities, and carry out extensive pre-release testing by internal and external experts, at multiple checkpoints along the way. Deployments of models will need to be controlled based on the assessed level of risk and evaluations of how well-placed users, regulators, and other stakeholders are to manage residual risks. Ongoing monitoring post-release will be essential to ensuring that guardrails are functioning as intended and that deployed models remain under human control at all times. In practice, we believe that the effective enforcement of such a regime will require us to go one

layer deeper in the tech stack to the AI datacenters on which highly capable AI models are developed and deployed.

- Much like the regulatory model for telecommunications network operators and critical infrastructure providers, we see a role for licensing providers of AI datacenters to ensure that they play their role responsibly and effectively to ensure the safe and secure development and deployment of highly capable AI models. To obtain a license, an AI datacenter operator would need to satisfy certain technical capabilities around cybersecurity, physical security, safety architecture, and potentially export control compliance. More specifically:

    - Operators of AI datacenters have a special role to play in securing highly capable AI models to protect them from malicious attacks and adversarial actors. This likely involves not just technical and organizational measures, but also an ongoing exchange of threat intelligence between the operator of the AI datacenter, the model developer, and a regulator.

    - Second, in certain instances, such as for scenarios that involve sensitive uses, the cloud operator on which the model is operating should apply the 'know-your-customer' principle – knowing the customers who are accessing the model. More thought and discussion will be needed to work through the details, especially when it comes to determining who should be responsible for collecting and maintaining specific customer data in different scenarios. The operators of AI datacenters that have implemented know-your-customer procedures can help regulators get comfortable that all appropriate licenses for model development and deployment have been obtained. One possible approach is that substantial uses of compute that are consistent with large training runs should be reported to a regulator for further investigation.

    - Third, as export control measures evolve, operators of AI datacenters could assist with the effective enforcement of those measures, including those that attach at the infrastructure and model layers of the tech stack.

    - Fourth, the AI infrastructure operator will have a critical role and obligation in applying safety protocols and ensuring that effective AI safety brakes are in place for AI systems that manage or control critical infrastructure. It will be important for the infrastructure operator to have the capability to intervene as a second and separate layer of protection, ensuring the public that these AI systems remain under human control.

For further detail on this licensing approach and how it fits into a broader regulatory framework that is aligned with the technology architecture for AI, please see our whitepaper, Governing AI: A Blueprint for the Future.

9. **Microsoft will support the development of an expanded 'know-your-customer' concept for AI services, building upon the same concept that has been implemented with respect to high-risk financial services.**

The “Know Your Customer” – or KYC – principle requires that financial institutions verify customer identities, establish risk profiles, and monitor transactions to help detect suspicious activity. It would make sense to take this principle and apply a KY3C approach that creates in the AI context certain obligations to know one’s cloud, one’s customers, and one’s content.

In the first instance, the developers of designated, powerful AI models must first “know the cloud” on which their models are developed and deployed. In addition, such as for scenarios that involve sensitive uses, the company that has a direct relationship with a customer, whether it be the model developer, application provider, or cloud operator on which the model is operating, should “know the customers” that are accessing it.

In addition, the public should be empowered to “know the content” that AI is creating through the use of a label or other mark informing people when something like a video or audio has been produced by an AI model rather than a human being. This labeling obligation should also protect the public from the alteration of original content and the creation of “deep fakes.” This will require the development of new laws, and there will be many important questions and details to address. But the health of democracy and future of civic discourse will benefit from thoughtful measures to deter the use of new technology to deceive or defraud the public. Microsoft will play its part by taking the following steps to adopt state-of-the-art provenance tools.

Increasing the trustworthiness of Microsoft AI systems

10. **Microsoft will commit that it will continue to ensure that our AI systems are designed to inform the public when it is interacting with an AI system and that the system's capabilities and limitations are communicated clearly, including via model- and application-level documentation.**

    Microsoft recognizes the role of documentation in helping our stakeholders understand the capabilities and limitations of AI models and applications and empowering them to make responsible use decisions. We will ensure that appropriate documentation is provided for AI models and systems that we release, and we will participate in efforts, such as the Partnership on AI's ABOUT ML initiative, to standardize that documentation across the industry.

    Microsoft will leverage system cards produced by OpenAI for OpenAI foundation models that Microsoft makes available via the Azure OpenAI Service. We will produce similar documentation for Microsoft trained foundation models, addressing their capabilities and limitations, system behavior, intended and unsupported use cases, and providing summarized results of red teaming and other evaluations. We will also provide information about best practices for improving system performance and integrating the core technology into systems built by our customers. Microsoft calls this type of product documentation Transparency Notes, and has made such documentation available for the Azure OpenAI Service and 14 of our Cognitive Services (see here).

    For AI-powered applications that we make available, such as the new Bing, Microsoft will continue to build AI systems designed to support informed decision making by the people who use them. We take a holistic approach to transparency which includes not only user interface features that inform people that they are interacting with an AI system but also educational materials, such as the new Bing primer, and detailed documentation of a system's capabilities and limitations, such as the Azure OpenAI Service Transparency Note. This documentation and these thoughtful experience design elements are meant to help people understand an AI system's intended uses and make informed decisions about their own use, consistent with the approach the White House has advocated for in its Blueprint for an AI Bill of Rights.

11. **Microsoft will increase investment in its academic research programs to ensure researchers outside Microsoft can access Microsoft's foundation models and the Azure OpenAI Service to undertake research and validate findings.**

    This expanded commitment builds on the success of our Turing Academic Program and Accelerating Foundation Models Research Program. It is designed to help the academic community gain API-based access to cutting-edge foundation models from Microsoft, as well as Microsoft's Azure OpenAI Service by which OpenAI models are made available. This will ensure that researchers can study frontier applications and the sociotechnical implications of these models. Microsoft will ensure that its program design accommodates API-based access by a diverse community of academic researchers, including researchers at Minority Serving Institutions across the United States.

    An important complement to providing such access is the development of governance best practices for the academic community engaged in frontier research on applications and the safety and security implications of highly capable models. Microsoft would welcome the opportunity to commit to supporting and collaborating with a multistakeholder group, including representatives across the academic community, to develop such practices.

12. **Microsoft will collaborate with the National Science Foundation to explore Microsoft's participation in a pilot project to inform efforts to stand up the National AI Research Resource, including by facilitating independent academic research relating to the safety of AI systems.**

Microsoft considers the establishment of the National AI Research Resource (NAIRR) to be of fundamental importance to the United States' leadership in AI innovation and risk mitigation. We are in discussions with the Technology, Innovation, and Partnership Directorate of the National Science Foundation to understand how Microsoft can contribute to a pilot project to validate the NAIRR concept and inform its broader implementation. We believe that this will advance multiple goals, including by facilitating academic research relating to the safety of AI systems.

We also would welcome and support an extension of the NAIRR to accommodate access by academic institutions in allied nations abroad, including the European Union, the United Kingdom, and Japan. A multilateral AI research resource would accelerate existing efforts to establish global norms and interoperable approaches to risk mitigation, including those underway in the U.S.-EU Trade and Technology Council and the G7.

13. **Microsoft will release an annual transparency report to inform the public about its policies, systems, progress, and performance in managing AI responsibly and safely.**

Transparency reports have proven to be an effective measure to drive corporate accountability and help members of the public better understand the state-of-the-art and progress toward goals. Microsoft believes transparency reports have a role to play in the responsible AI context too, and so we will release an annual transparency report to inform the public about our policies, systems, progress, and performance in managing AI responsibly and safely. If adopted across the industry, transparency reports would be a helpful mechanism for recording the maturing practice of responsible AI and charting the cross-industry progress made.

Specifically, Microsoft's annual transparency report will address the functioning and ongoing development of our governance systems, in addition to providing case studies on the implementation of responsible AI measures.

As further explained in commitment , Microsoft will also continue to produce product-specific documentation, such as the Transparency Notes we have published that describe the capabilities, limitations, and intended use cases for the core AI technology that we make available via APIs and that our customers build upon. An example of such documentation is the Azure OpenAI Service Transparency Note; similar documentation for 14 of our Cognitive Services is available here.

14. **Microsoft will support the development of a national registry of high-risk AI systems that is open for inspection so that members of the public can learn where and how those systems are in use.**

    Public trust in AI systems can be enhanced by demystifying where and how they are in use. For high-risk AI systems, Microsoft supports the development of a national registry, maintained by the Federal Government and accessible via a website, that would allow members of the public to review an overview of the system as deployed and the measures taken to ensure the safe and rights-respecting performance of the system.

    For this information to be useful for members of the public, it should be expressed at the system level, provide details about the context of use, and be written with non-technical audiences in mind. To achieve this end, the United States could implement an approach that draws from an effective European model: several European cities have adopted the Algorithmic Transparency Standard and developed accessible explanations of how they use AI (see, for example, the City of Amsterdam's Algorithm Register). Microsoft would be pleased to contribute to a U.S. Government-led effort to create a template for the information to be included in the high-risk registry. We would also support sharing any finalized templates with our customers that are deploying high-risk AI systems to help promote the national registry and the transparency function it performs.